

8 Drittes Moment: Good Matches

Digitale Personendaten haben einen doppelten Verweisungshorizont. Erstens verweisen sie auf das, was sie als Spuren »präsentieren« (Krämer et al., 2007): auf individuell und sozial bedeutsame Sinngehalte und Interaktionen. Durch Techniken des »encoding« und der Quantifizierung oder Vektorisierung werden Verhaltensweisen von Nutzerinnen abgeflacht und von Sinnüberschüssen befreit. Welche Intentionen oder bedeutungsvollen Interaktionen mit anderen Nutzerinnen, Gruppen oder Dingen verbunden waren, spielt formal keine Rolle mehr – auch wenn das in der Beurteilung der Daten weiterhin relevant ist. Diese »formal indifference« macht singuläre Verhaltensweisen zu vergleichbaren Verhaltensdaten, zu einer Ressource, mit der gerechnet werden kann.

Zweitens verweisen Personendaten prospektiv auf etwas, was in den Daten steckt. Etwas, das nicht explizit mitgeteilt wurde, aber aus den Daten »abgeleitet« werden kann. Wie im Beispiel der »schönen Daten«, ist Earlybird nicht primär daran interessiert, wohin Teilnehmerinnen reisen und was sie dort tun möchten, sondern daran, was sich sonst noch mit den Daten machen lässt.¹

Das Wertversprechen besteht im »impliziten« Gehalt der Wettbewerbsantworten: den kategorialen Zugehörigkeiten, die sich in den Daten ausdrücken, und den Rückschlüssen auf zukünftige oder wahrscheinliche Verhaltensweisen oder Präferenzen, die daraus gezogen werden können.

Wie das Beispiel von Earlybird, aber auch zahlreiche andere Fälle klar machen, besteht dieser »sekundäre« Zweck vor allem im »social sorting«, d.h. in der Ableitung von kategorialen Zugehörigkeiten wie Geschlecht (Cheney-Lippold, 2011; Buolamwini & Gebru, 2018), sexueller Orientierung (Wang & Kosinski, 2018) für Marketing wertvolle Hinweise auf Lebensereignisse wie

1 Siehe Beauvisage & Mellet (2020) zu »repurposing« von Personendaten oder Mayer-Schönberger & Cukier (2013, S. 173) zur Zweitverwendung von Daten: »In the era of big data [...] much of data's value is in secondary uses that may have been unimagined when the data was collected«.

Umzüge oder Schwangerschaften (Ebeling, 2016), Konsumneigungen (Markenzie, 2018) oder Credit-Scores (Fourcade & Healy, 2017b).

The surveillance system obtains personal and group data in order to classify people and populations according to varying criteria, to determine who should be targeted for special treatment, suspicion, eligibility, inclusion, access, and so on. [...] It sieves and sorts for the purpose of assessment, of judgement. (Lyon 2003, S. 20, siehe auch Gandy 1993)

Marktakteure verwenden algorithmische Systeme der Kategorisierung und der Bewertung (Fourcade & Healy, 2016), um ihre Kundschaft zu differenzieren (Zwick & Denegri Knott, 2009; Turow, 2003). Entgegen der Annahme von Ökonomen ist der Markt nicht blind für sozialen Status: »markets see social differences very well and thrive on them. Like states, market technologies make societies more legible« (Fourcade & Healy, 2016, S. 562). Solche Kategorisierungen sind für Individuen folgenreich, da sie bestehende soziale Differenzierungen und Ungleichheiten aufgreifen und neue Differenzierungen erzeugen. Sie resultieren in »classification situations«, die Lebenschancen beeinflussen. Anstrengungen mit dem Ziel »to know your customer« – in der Sprache des Marketings – führen auf der Seite der getrackten, profilierten, kategorisierten und gescorten Individuen zu einer neuen Form von Kapital: »übercapital«, »a form of capital arising from one's position and trajectory according to various scoring, grading and ranking methods« (Fourcade & Healy, 2017b, S. 14).

Insbesondere Credit-Scores sind (zumindest im US-amerikanischen Kontext) ein zentraler Generator von »classification situations«, d.h. der Position im Kredit-Markt und der Art möglicher Kredite. Auch darüber hinaus kommen Credit-Scores in nicht kreditbezogenen Kontexten wie beispielsweise bei Job-Bewerbungen zur Anwendung (siehe Rona-Tas 2017 zu »off-label use« von Credit-Scores).

Auf Vatin (2013) referierend deuten Fourcade & Healy (2016) an, dass die Bewertung (»valuation«) von Individuen von der Verwertung (»valorization« bei Vatin, »value-extraction« bei Fourcade und Healy) teilweise entkoppelt sein kann. Ihr Beispiel der Credit-Scores verdeutlicht dies: Geringes Überkapital im Sinne eines niedrigen Scores ist nicht automatisch mit einem geringen Marktwert verbunden.

Someone with very low übercapital may actually be very valuable from the point of view of their classification in the market. For instance, companies

may pay dearly to acquire lists of people with gambling problems, or chronic diseases requiring medication. (Fourcade & Healy, 2016, S. 14)

Es ist dementsprechend nicht alleine die Höhe des Scores oder die Menge an Überkapital, die für die In-Wert-Setzung eines individuellen Datenprofils ausschlaggebend ist. Es interessieren nicht nur die Spitzenplätze in den Scorings. Nicht nur die bestplatzierten, sondern auch Personen mit einem »subprime« Credit-Score erhalten Kredite – nur eben zu schlechteren Konditionen. Es gibt ein »Kreditprodukt« für jeden »Kredittyp«.

Die Nutzung digitaler (und zum Teil analoger) Güter und Dienstleistungen generiert Daten, welche die Nutzerinnen hinter ihrem Rücken als kreditwürdig, »subprime« oder irgendwo dazwischen »markieren« (Fourcade & Healy, 2016, S. 17). Unternehmen nutzen diese Daten, um »good matches« zwischen Konsumentinnen und Produkten herzustellen – zum Beispiel zwischen einer wohlhabenden Kundin und einer Kreditkarte mit hoher Limite. Im Kern geht es darum, potenzielle Kundinnen mit denjenigen Angeboten zu adressieren, welche die höchste Annahme- und Profitwahrscheinlichkeit haben.

Fourcade und Healy schlagen vor, Viviana Zelizers (1994; 2011) Konzept der »good matches« auf digitale Daten anzuwenden. Wie Zelizer darlegt, markieren Formen des Zahlens Beziehungen zwischen Geberinnen und Empfängerinnen. Mit »good matches« bezeichnet sie die Übereinstimmung von Geld- und Bezahlformen mit Kategorien von Personen oder Beziehungen. Fourcade erläutert dies anschaulich am Beispiel von Studentinnen:

College students worry tremendously about making mistakes when offering in-kind or in-cash gifts to their less privileged classmates, for instance. Is their generosity over-the-top and patronizing? Is the medium appropriate? And if they are on the receiving end, should they accept it and how? Will the gift put them in debt, and undermine the presumed equality in the relationship? The students seem to be aware, in a confused sort of way, that the delicate equilibrium that their cross-class friendships have achieved may be at stake in such exchanges. At stake, too, are their feelings about themselves, their social persona, their place in the broader Princeton society. (Fourcade, 2016a)

Die richtige Art von Transfer (z.B. Geschenke), kombiniert mit der richtigen Form (z.B. Cash), muss mit der spezifischen Art der Beziehung in Abstimmung gebracht werden. Für Studentinnen stehen dabei die Beziehungen

zu den anderen Studentinnen auf dem Spiel. Auch für die von Tsing beschriebenen Matsutake-Händler stellt sich ein ähnliches Problem, wenn sie ihre Pilze mit ganz bestimmten Abnehmerinnen im Kopf sortieren, um die richtige Käuferin mit den richtigen Pilzen zusammenzubringen und so die Beziehung zu erhalten. »Good matches« »[get] the economic work of the relationship done and [sustain] the relationship« (Zelizer, 2006, S. 307).

Karen Levy liefert ein weiteres Beispiel, indem sie das Konzept auf Trackingdaten anwendet: »people constitute and enact their relations with one another through the use and exchange of data« (Levy, 2013, S. 75). Das verdeutlicht sie am Beispiel einer Smartphone-App, die es Nutzerinnen erlaubt, den Standort von Freunden in Echtzeit zu verfolgen. Die Beziehung von zwei Personen wird definiert darüber, ob sie sich von ihrem abwesenden Gegenüber tracken lassen. Damit differenzieren die Überwacherinnen zwischen jenen Freundinnen, die sie überwachen, und jenen, die sie nicht tracken beziehungsweise dies nicht zulassen.

Levy belässt es bei der Untersuchung, wie über Daten persönliche Beziehungen zwischen Individuen definiert und aufrechterhalten werden. Wie Fourcade & Healy (2016) anregen, lassen sich aber auch kommerzielle Beziehungen hinsichtlich der Frage untersuchen, wie Unternehmen über Praktiken des Trackings und der Datenverarbeitung ökonomische Beziehungen erzeugen und erhalten. In diesem Sinne sind (ökonomische) Transaktionen nicht nur das Resultat bestehender Beziehungen, sondern »the exchange itself produces varied forms of ties, in the sense of »connections« between various entities« (Cochoy et al., 2017, S. 23), »between persons, between things, and between persons and things« (Cochoy et al., 2017, S. 27). Für Unternehmen der digitalen Ökonomie stehen analog zu den Studentinnen oder den Matsutake-Händlerinnen die guten Beziehungen zu ihren Nutzerinnen (und ihren Werbekundinnen) auf dem Spiel.

Im Kern von »good matches« der digitalen Ökonomie steht die Herstellung eines Passungsverhältnisses zwischen Nutzerin und Dingen, das »engagement« aufrecht erhält (Seaver, 2018) und weiteren Konsum motiviert. Den richtigen Film, das richtige Produkt, den richtigen News-Artikel, den richtigen Post, die richtige Werbung, den richtigen Song als Nächstes zu präsentieren, soll die Nutzerin zum Klick (oder zum Kauf) veranlassen, sie länger auf der Seite halten und so die Beziehung zur Plattform (i.e. Netflix, Instagram, Spotify, etc.) festigen.

Die Herstellung von »good matches« lässt sich dann als eine organisatorische Praxis verstehen, in der Personendaten für die Unternehmen der di-

gitalen Ökonomie in Wert gesetzt werden (Kallinikos & Alaimo, 2019). Über Personendaten lassen sich Nutzerinnen und Dinge sowie Nutzerinnen und Dinge je untereinander in Beziehung setzen und daraus neue Relationen ableiten. »Good matches« sind die Grundlage eines (reziproken) Beziehungsgeflechts zwischen Unternehmen, Nutzerinnen, Dingen und je nach dem auch Werbekundinnen und Entwicklerinnen (Fourcade & Kluttz, 2020).

8.1 Relationen ableiten

Im ersten Moment der Datafizierung werden Nutzerinnen und Unternehmen miteinander verwickelt, um Personendaten zu erzeugen. Im zweiten Moment werden Relationen und Bedeutungen entfernt oder operativ ignoriert, damit die Verhaltensweisen zu Daten werden, mit denen sich rechnen lässt. Im dritten Moment wird gerechnet, um aus den Personendaten neue Relationen beziehungsweise »good matches« zwischen Nutzerinnen und Dingen abzuleiten. Im Idealfall – d. h. in funktionierenden Daten- und Empfehlungsinfrastrukturen – geschehen alle drei Momente mehr oder weniger gleichzeitig, so dass ein Klick zum nächsten führt. Die Relation zwischen Nutzerin und Unternehmen wird aufgefrischt, neue Daten den Profilen hinzugefügt und daraus werden neue Relationen abgeleitet, die zu weiteren Klicks motivieren.

Das ist der Idealfall. Aber wie wird die Lücke zwischen Nutzerin und dem nächsten »Ding« überbrückt? Wie werden »good matches« hergestellt?

Die kurze Antwort: Es sind Algorithmen, welche die Lücke zwischen Nutzerin und Dingen durch »inferring across the gaps« schließen (Amoore, 2011, S. 27). Die »Magie« der Algorithmen (Ziewitz, 2016) ermöglicht »inductive leaps« (Gregg, 2015), um aus Personendaten weitere Daten abzuleiten (Amoore, 2011), die über Passungsverhältnisse von Nutzerinnen und Dingen Auskunft geben.

Um dieses dritte Moment der Datafizierung im Detail zu verstehen, wechseln wir von einer »ontologischen« Perspektive auf Personendaten zu einer »epistemologischen« (Mai, 2016). Unser Fokus verschiebt sich vom ersten Verweisungshorizont darüber, was Daten repräsentieren, auf den zweiten: was aus Daten abgeleitet werden kann.

Data collection is ontologically oriented; it focuses on data as representing facts about states of affairs in the world: people and activities and the inter-

relation between places, times, other people, activities, and intentions. Data processing and analysis is epistemologically oriented; it focuses on the facts or realities that data can generate as they are processed and analyzed. (Mai, 2016, S. 198)

Mai illustriert die Differenz anhand eines Problems in der Debatte um Privatsphäre: Er argumentiert, dass Privatsphäre nicht nur durch die *Sammlung* digitaler Personendaten verletzt werde, sondern vor allem durch deren *Verarbeitung*. Er erläutert die Differenz am berühmten Beispiel von Targets »pregnancy prediction score« (Duhigg, 2012). Dieser veranlasste Target, einer jungen Frau Werbung für Schwangerschaftsprodukte zu senden und machte dabei ihren Vater auf die Schwangerschaft aufmerksam.²

The woman's privacy was not violated due to the collection of data (she had presumably volunteered the information), but it could be argued that her privacy was violated due to data processing and analysis. (Mai, 2016, S. 198)

Mai schlägt ein »datafication model of informational privacy« vor, das den Fokus von der Sammlung auf die *Verarbeitung* von Daten verschiebt: Dies erlaubt es zu erkennen, dass nicht die, möglicherweise mit Einverständnis gegebenen, Daten das Problem sind, sondern neue Informationen und Relationen, die daraus »abgeleitet« werden können.

Das singuläre Datum, »junge Frau kauft unparfümierte Lotion«, wird in diesem Sinne erst bedeutungsvoll, wenn sich dieses an sich bedeutungslose Datum zu anderen Daten in Beziehung setzen lässt. Dass sie zum Beispiel Vitaminzusätze und extragroße Wattebäusche gekauft hat und dass diese Produkte von Frauen gekauft wurden, die sich in Targets »Baby Shower«-Programm eingeschrieben hatten (Duhigg, 2012).

The individual datum produced from a single user at a given time and place (e.g. posting a picture of a meal to Instagram) is both nearly meaningless [...] and valueless [...] until it is linked to the user's past data produced, the user's network of other users, the user's growing network of location data, and the temporal rhythms and spatial patterns embedded in data from many users. (Thatcher et al., 2016, S. 995)

2 Ein oft verwendetes Beispiel, das wie kein anderes zugleich für die »power of the algorithm« (Beer, 2017) steht (zum Beispiel im Buch »Data Science for Business« von Provost & Fawcett 2013) und für die Gefahr der Überwachung und des Data Mining (siehe zum Beispiel: Marwick 2014).

Der Versuch, Personendaten zu individuellen »assets« der Nutzerinnen zu machen (siehe prominent: Lanier 2014), scheitert daran, dass Personendaten nicht schon an sich als Güter betrachtet werden können (Beauvisage & Mellet, 2020, S. 84). Personendaten entfalten ihren Wert erst – beziehungsweise werden erst zur »digital commodity« –, wenn sie mit anderen Daten in Verbindung gesetzt werden (siehe auch: Charitsis et al. 2018).³

Ich zeige zwei unterschiedliche Arten, wie Daten miteinander verlinkt werden, um neue Informationen und prospektive Relationen daraus abzuleiten. Die erste Form der Verlinkung nennen Lee et al. (2019) »folding«: Dabei werden Datensätze, Klassifikationen und Normalitäten unterschiedlichen Ursprungs miteinander verbunden, um neue Ordnungsweisen und Perspektiven auf die Welt hervorzubringen. Die zweite Form, die ich im Anschluss an Bettina Heintz' (2010; 2016) Vergleichssoziologie »Relationierung« nenne, bezeichnet die unzähligen algorithmischen Operationen des Vergleichs, in denen Nutzerinnen und/oder Dinge paarweise zueinander in Relation gesetzt werden, um für spezifische Nutzerinnen passende Dinge zu eruiieren.

8.2 Faltungen

Eine wachsende sozialwissenschaftliche Literatur befasst sich mit den Folgen weitgehender Profilierungs- und Kategorisierungsmaßnahmen privatwirtschaftlicher und öffentlicher Akteure. Sie kritisiert Verzerrungen in Daten und Algorithmen, die zu ungerechten Resultaten führen – entgegen den Versprechungen, dass Algorithmen vermeintlich neutraler und weniger vorurteilsbehaftet entscheiden könnten. Wie sich zeigt, reproduzieren Algorithmen tendenziell soziale Ungleichheiten. Lee et al. (2019) weisen auf eine implizite Annahme solcher Ansätze hin: Wenn Algorithmen bloß »richtig« funktionieren würden, würden sie auch objektive und faire Resultate produzieren. Diese »objektivistische« Perspektive auf Algorithmen ist sowohl kritisch als auch affirmativ. Anstatt zu fragen, inwiefern Algorithmen oder

3 Beauvisage & Mellet (2020, S. 91) weisen auch darauf hin, dass nicht Personendaten gehandelt werden – »it appears that personal data are not marketized as »the product« in this market, although they came to constitute a key element in its architecture« –, sondern Tools und Dienstleistungen, mit denen sich Personendaten nutzen lassen. Dazu später mehr.

Daten fehlerhaft sind – und damit indirekt: wie sie sich verbessern oder korrigieren lassen, um »korrekte« Resultate zu produzieren – schlagen Lee et al. (2019) eine alternative analytische Perspektive auf Algorithmen vor. Sie betrachten Algorithmen als Teil von soziotechnischen Systemen, die durch »operations of folding« neue Ordnungen und Relationen generieren:

an analytical approach focusing on folding – on relating things that were previously unconnected – is better able to account for the varied processes by which algorithms order society and nature. (Lee et al., 2019, S. 2)

Lee et al. (2019) beschreiben drei Möglichkeiten, wie Algorithmen als soziotechnische Systeme verschiedene Einheiten in eine Relation bringen können, zwischen denen zuvor keine Relation bestanden hatte. Die Operation der »Proximierung« demonstrieren sie an einem Re-Mapping der Aids-Epidemie von sozialen Kategorien wie »homosexuell« oder »Heroin-süchtig« in räumliche Distanzen der Ausbreitung: »a set of social proximities was replaced with a set of spatial proximities« (Lee et al., 2019, S. 4). »Universalisierung« macht aus einem Flickenteppich an heterogenen, unter spezifischen Umständen erzeugten, Datensätzen eine universelle Tatsache. Als Beispiel verwenden sie eine Karte der Zika-Verbreitung, welche die Schwierigkeiten und Unsicherheiten der Datenerhebung unsichtbar machen. »Normalisierung« beschreibt, wie spezifische Normalitätsannahmen in Algorithmen »eingefaltet« werden. Algorithmen zur Vorhersage von Finanzmarktentwicklungen, die auf einer Glockenkurve basieren, gehen von anderen Normalitätsannahmen aus als solche, die sich auf die Power-Law-Verteilung stützen. Im ersten Fall wird angenommen, dass kleine Bewegungen normal und Krisen selten sind. Die Power-Law-Verteilung geht hingegen davon aus, dass große Veränderungen (i.e. Krisen) weitaus öfter vorkommen als die Glockenkurve suggeriert (Lee et al., 2019, S. 7). Ihr Fazit: Algorithmen bilden nicht eine soziale Realität einfach ab. Sie schaffen durch Faltungen neue Ordnungen, in denen sie Relationen zwischen verschiedenen Einheiten, Zeiten und Räumen bilden.

Rather than thinking about objects, relations and concepts as stable entities with fixed distances and properties, we might attend to how different topologies produce different nearnesses and rifts. In this way, technologies, such as algorithms, can be understood as folding time and space as much as social, political and economic relations [...]. By analysing algorithms in this manner, we argue that we can gain a better understanding of how they become part of ordering the world: sometimes superimposing things that

might seem distant and sometimes tearing apart things that might seem close. (Lee et al., 2019, S. 3).

»Folding« scheint mir eine geeignete Metapher zu sein, um zu beschreiben, wie und auf welcher Grundlage Algorithmen operieren. Wie Roberge und Melançon betonen, haben Algorithmen eine »double agency«: »they construct meanings as much as they are shaped by meanings« (2017, S. 308). Der Begriff des Faltens bezieht sich darauf, welche bestehenden Kategorien, Deutungsweisen und Selbstverständlichkeiten in algorithmische Systeme »eingefaltet« werden. Was Fourcade und Healy für das Verhältnis von Quantifizierung und Kategorisierung feststellen, lässt sich auf algorithmische Systeme ausweiten:

Quantification not only implies classification, it implies classifications on top of other classifications – indeed a classificatory architecture that pulls in variegated ways of boxing and measuring people and things to some end. The pristine numerical output of a final score may bear a tangled relationship to its underlying strata of classes, groups, and types. (Fourcade & Healy, 2017a, S. 289)

Die Idee der Faltung betont, dass bei jeder weiteren Faltung bestimmte Kategorien unsichtbar werden, ohne gänzlich zu verschwinden: Sie werden selbstverständlich und können möglicherweise auch wieder mobilisiert werden.

Das folgende Beispiel von Earlybirds Kategorisierungsexperiment zeigt, wie so unterschiedliche Dinge wie Wettbewerbsantworten, die deutschsprachige Wikipedia, Marketingsegmente, der TF-IDF-»Algorithmus« und Expertenurteile ineinander gefaltet werden, um Nutzerinnen »algorithmisch Kategorien zuordnen« zu können. Aus extrem schwierigen qualitativen Urteilen darüber, welche Nutzerinnen zu welchen Kategorien gehören, werden berechenbare Distanzen, die darüber Auskunft geben, wie nahe sich Teilnehmerinnen und Kategorien im abstrakten, semantischen Raum von Wikipedia sind. Das Experiment zeigt auch, dass menschliche (Experten-)Urteile nicht überflüssig werden, sondern mit eingefaltet werden müssen. Am Ende steht Earlybird vor sauber kategorisierten Teilnehmerinnen, denen nichts mehr von Wikipedia, Experten oder Wettbewerben anzusehen ist.

Falten und Vergleichen

Erste Faltung: Wikipedia

Wir haben bereits gesehen, dass Earlybirds Wettbewerbsteilnehmer erst dann vergleichbar werden, wenn die Wettbewerbsantworten in einem gemeinsamen Kontext positioniert werden können: Das ist der von Wikipedia aufgespannte Vektorraum. Der Kontext des Reisewettbewerbs in all seinen Facetten (Formulierung der Frage, Darstellung der Benutzeroberfläche, Rekrutierung der Teilnehmerinnen, Preise, Hoffnungen der Teilnehmerinnen, etc.) wird für die weiteren Verarbeitungsschritte unsichtbar gemacht und durch den »Kontext Wikipedia« ersetzt. Der Begriff des algorithmischen Falten macht uns darauf aufmerksam, dass die mathematische Transformation von Text keine einfache De-Indexikalisierung oder Reinigung von Kontext ist: Es resultiert kein objektiver Blick von Nirgendwo (Haraway, 1988). Wie Michael Lynch (1992) in einer Auseinandersetzung mit dem ethnomethodologischen Begriff der Indexikalität beobachtet, lässt sich Indexikalität nicht »reparieren«. Missverständnisse und Unklarheiten lassen sich nur durch einen Bezug auf weitere indexikalische Äußerungen »reparieren«. Die Formalisierung indexikalischer Äußerungen, deren Bedeutungen immer auch im Ungesagten, Unsagbaren und schon Vorausgesetzten besteht, löst das Problem der Indexikalität, indem die Äußerung in einen neuen, formalisierten Kontext (hier: Wikipedia) gestellt wird.⁴

Wie Urs, Simons Assistent an der Hochschule, ausführt, verwenden sie dazu die Methode der »explicit semantic analysis« (ESA) von Evgeniy Gabrilovich und Shaul Markovitch (2007). In ihrem Paper »Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis« schlagen die beiden Autoren ESA als eine neue Methode vor, um maschinelles Textverständnis dem menschlichen näher zu bringen:

[H]umans interpret the specific wording of a document in the much larger context of their background knowledge and experience. [...] in order to process natural language, computers require access to vast amounts of common-sense and domain-specific world knowledge. (Gabrilovich & Markovitch, 2007, S. 1606)

Was ein Mensch versteht, wenn er oder sie das Wort »Wald« hört, ist von individuellen und kollektiven Wissensbeständen und Erfahrungen abhängig. Das soll im Fall von ESA durch Wikipedia simuliert werden. Die Methode ba-

siert auf einer Indexierung von Wikipedia, d.h. es wird eine Tabelle erstellt, die in den Zeilen alle vorkommenden Wörter und in den Spalten alle Artikel (»concepts«) enthalten. In den Schnittpunkten von Wörtern und Artikeln werden Gewichtungen eingetragen, die darüber Auskunft geben, ob ein Wort vorkommt und wie relevant es im betreffenden Artikel ist. Wörter, die im betreffenden Artikel oft vorkommen, im Gesamtkorpus aber eher selten sind, erhalten ein höheres Gewicht (TF-IDF): »These weights quantify the strength of association between words and concepts« (Gabrilovich & Markovitch, 2007, S. 1607).⁵ Wie bereits in Kapitel 7 beschrieben, wird durch diese Tabelle ein Vektorraum mit potenziell so vielen Dimensionen, wie Wikipedia Artikel hat, aufgespannt. Für die deutschsprachige Wikipedia sind das 2,5 Millionen. Die Bedeutung eines interessierenden Worts lässt sich bestimmen, indem dieses Wort in der Tabelle nachgeschlagen wird: Wo im Vektorraum befindet sich dieses Wort und welche anderen Wörter sind am nächsten? Relationen der Ähnlichkeit und Differenz beziehungsweise Nähe und Distanz bestimmen seine Bedeutung (siehe die Grafiken 5 und 6 im vorhergehenden Kapitel).

Als größte und mittlerweile auch anerkannte Enzyklopädie betrachten die Autoren Wikipedia als geeignetes Korpus, um die maschinelle Repräsentation von Textdaten zu verbessern. Wikipedia-Artikel eignen sich insbesondere deshalb, weil es sich dabei um »natural concepts«, »defined by humans themselves« (Gabrilovich & Markovitch, 2007, S. 1606) handle und sich jeder Artikel mit nur einem spezifischen Thema im Detail befasse. Eine zentrale Rolle spielt dabei die Idee der »semantic relatedness«, die den opaken Prozess der menschlichen Interpretation approximieren soll.⁶ »[B]ackground knowledge and experience«, das in den Köpfen der Interpreten steckt, wird über Wikipedia-»concepts« operationalisiert. Wikipedia, wie auch andere Enzyklopädien, scheinen die Funktion als neutrales universelles Korpus einzunehmen, in dem das Wissen und die Erfahrungen aller gleichermaßen stecken.⁷ Doch Simon und sein Team ziehen in Erwägung, dass Wikipedia möglicherweise nicht der richtige »Kontext« für die Kategorisierung der Earlybird-Member darstelle. Sie stellen in Aussicht, das Archiv aller »20 Minuten«-Artikel zu verwenden, falls sie Zugang dazu erhalten.⁸ »20 Minuten« sei näher an der Jugendsprache als Wikipedia und sei für den Schweizer Kontext angemessener, da darin auch spezifische Schweizer Begriffe wie »Velo« und »Glacé« abgebildet seien. »Semantic relatedness« ist davon abhängig, welcher »Kontext« in den Kategorisierungsalgorithmus »eingefaltet« wird.

Zweite Faltung: Marketingsegmente

Wie Gabrilovich und Markovitch zeigen, lässt sich die Ähnlichkeit beliebiger Wörter oder auch ganzer Texte im Kontext von Wikipedia eruieren. Earlybird ist aber nicht an irgendwelchen Wörtern interessiert. Ihnen geht es darum, die Freitextantworten der Teilnehmerinnen mit Beschreibungen von Marketingsegmenten in Beziehung zu setzen, um daraus automatisch kategoriale Zugehörigkeiten der Member abzuleiten. Simon lässt die Milieubeschreibungen, die er von Earlybird erhalten hat, von einer Kollegin an der Fachhochschule mit weiteren beschreibenden Begriffen »anreichern«.

Die Definitionen der Marketingsegmente unterliegt im Verlauf des Kategorisierungsexperiments einem starken Wandel, wie ich erfahre. Die auf 20 Schlagwörter angereicherten Beschreibungen werden weiter optimiert, so dass die Milieus zu einem späteren Zeitpunkt nur noch durch drei bis sieben Schlagwörter charakterisiert sind. Gemäß Simon und Urs hätten sie mit diesen kürzeren, dafür aber »präziseren« Schlagwortlisten die besten Resultate erzielt.

Die automatisierte beziehungsweise »algorithmische Lösung« – das heiße: ein »Computerprogramm« zu schreiben, das »Hypothesen« aufstellen kann, welche Member zu welchen Zielgruppen gehören – für das Zuordnungsproblem ist eine spezifische Variante des Vergleichs: Erst werden die unterschiedlichen Einheiten paarweise zueinander in Beziehung gesetzt, um dann diese Beziehungen zu vergleichen (siehe genauer in den Kapiteln 8.3 und 10).

Dazu werden sowohl die Milieubeschreibungen als auch die Freitextantworten in den Vektorraum übertragen. Aus den Vektoren der einzelnen Wörter eines Milieus wird der »Centroid« berechnet, d.h. das geometrische Zentrum aller Wörter eines Milieus. Das Gleiche wird für die Freitextantworten gemacht. In Pseudo-Code ausgedrückt lautet die Vorgehensweise wie folgt:

1. Generiere die ESA-Datentabellen aus Wikipedia (Wörter in den Zeilen, Wikipedia-Artikel in den Spalten, TF-IDF-Werte in den Feldern).
2. Bestimme die Wörter der Marketingsegmente; bestimme ihre Lage im Vektorraum; berechne den Centroiden für jedes Segment.
3. Für jede Freitextantwort:
 - (a) Bestimme die Lage aller Wörter im Vektorraum; berechne den Centroiden.

- (b) Berechne die Distanz zu jedem Segment.
- (c) Ordne die Segmente aufsteigend nach dem Kriterium »Distanz zur Antwort« und retourniere jenes Segment mit der niedrigsten Distanz.

Segmente und Antworten werden vergleichbar gemacht, indem sie in einem gemeinsamen Vektorraum positioniert werden. Differenzen und Ähnlichkeiten lassen sich quantitativ als Distanzen ausdrücken. Wie wird nun verglichen? Die Spezifik des algorithmischen Vergleichs besteht darin, dass Segmente und Antworten paarweise *relationiert* werden. Das heißt, die Relationen aller Antwort-Segment-Kombinationen werden verglichen, ausgedrückt als Maß ihrer »semantic relatedness« (Gabrilovich & Markovitch, 2007). Es interessiert nicht, welche Antwort den höchsten Hedonisten-Score aufweist. Entscheidend ist, für welches Segment eine spezifische Antwort den höchsten Score erhält.

Die Zuordnung von Segmenten operiert als Vergleich: Der Vergleich von paarweisen Antwort-Segment-Relationen produziert kategoriale Relationen im Sinn von »nominal judgements« (Fourcade, 2016b). In der Herstellung handelt es sich nicht um absolute, sondern unscharfe, kontinuierliche Zuordnungen: Jede Antwort erhält für jedes Segment einen Score, der darüber Auskunft gibt, wie gut das Milieu past.⁹ Aus diesen kontinuierlichen Scores werden nominale Kategorien: »Continuous measures are cut into ranked scales, which in turn come to life as classes or categories of person, organization, or group« (Fourcade & Healy, 2017a, S. 287). Wer für das Segment »Hedonist« den höchsten Score aufweist, wird in den anschließenden Marketingkampagnen als Hedonist behandelt.¹⁰

Dritte Faltung: Der Goldstandard, oder: händische Kategorisierung durch Expertinnen

Die Frage sei nun: »Wie messen wir, ob das funktioniert?« Woher weiss Earlybird, ob es sich nicht einfach um einen »mehrbesseren Zufallsgenerator« handle? Für Simon und sein Team sei klar, dass sie das nicht selber entscheiden könnten: Sie müssten die Vorgehensweise mit jener einer Person vergleichen, die beurteilen kann, was zusammengehört: »das unterstellen wir euch«. Mit »euch« meint er mich, Sabina und Anna. Sabina ist die Marketingleiterin, Anna ist die Mitarbeiterin eines befreundeten Jugendmarketingbüros, wo

sie »content creation« und »storytelling« macht. Sie war bei der Durchführung des Wettbewerbes für das Reisebüro ebenfalls involviert. Wir wurden einberufen, um einen »Goldstandard« zu erarbeiten, der als »benchmark« für den Algorithmus dienen kann, d.h. es soll dann verglichen werden können, wie gut der Algorithmus darin ist, Milieus mit TeilnehmerInnen zu verbinden. Gemäß Simon brauche es dazu erfahrene Marketingexperten – mir als Soziologe wird ebenfalls die notwendige Expertise zugesprochen – um die sechs verschiedenen Milieus den TeilnehmerInnen zuzuordnen.¹¹

Simon bereitete eine Excel-Datei für uns vor, in der wir alle Wettbewerbsantworten vorfanden, die jeweils um zwei Spalten ergänzt waren. In die erste Spalte sollten wir dasjenige Milieu eintragen, von dem wir »vermuten«, dass es am besten passt. Die zweite Spalte sei für ein zweites Milieu reserviert, das nicht so gut wie das erste passe. Wir könnten auch angeben, dass keine Zuordnung möglich sei. Sie würden dann eine »Konsensdiagnose« vornehmen: Die erste Priorität erhalte zwei Punkte, die zweite Priorität einen Punkt. Unsere Resultate könnten sie dann zu einem »Konsens« »konsolidieren«. Es lasse sich dann noch filtern: Wenn drei Personen an einer Stelle das Gleiche sagen, sei die »reliability« hoch. Wenn alle etwas Unterschiedliches sagen, könne man den Datensatz verwerfen. Wenn wir alle unterschiedlicher Meinung seien, könne man vom Algorithmus nicht erwarten, dass er das Richtige ausgeben. Jene Fälle, in denen wir als Experten und ExpertInnen übereinstimmen, verwenden sie als »Goldstandard« – als »absolute Wahrheit«, an dem sich der Algorithmus messen muss.

War im Email und im Gespräch mit Beni noch die Rede davon, dass Sabina, Anna und ich gemeinsam ausdiskutieren sollten, welchen TeilnehmerInnen wir welche Segmente zuordnen, so ist jetzt gefordert, dass wir dies unabhängig tun. Es scheint so zu sein, als wäre die »intelligence of the crowd« nicht deliberativ, sondern die Summe aller Einzelurteile. Anna fragt nach, wie sie damit umgehen soll, wenn jemand die Schweiz als Reiseland angebe – was eher bei traditionell orientierten Leuten beliebt sei – und dann aber etwas sehr Spezielles dazu schreibe. Simon antwortet, dass wir das »frei aus dem Bauchgefühl« machen sollten, aber alle Länder und alle Antworten berücksichtigen sollten: Wir sollen auf »das Ganze« schauen. Die Antworten seien auch geordnet: Was als erstes angegeben wird, sei wichtiger.

Wir einigen uns darauf, in einem Monat je die ersten 600 Wettbewerbsantworten zu kategorisieren. Ich erhalte die Rolle des Koordinators. Tags dar-

aufschickt mir Simon den Datensatz. Beim Durchsehen der ersten 50 Antworten fällt mir auf, dass Länder und Freitextantworten verschoben sind. Zudem sind einige Textantworten italienisch und französisch. Ich kontaktiere Simon: Er ist froh, dass ich ihn auf den Fehler aufmerksam mache und schickt mir die richtige Version des Datensatzes. Die italienischen und französischen Textantworten entferne ich, bevor ich Anna und Sabina den Datensatz zur individuellen Kategorisierung schicke.

Die Kategorisierung einzelner Wettbewerbsantworten erweist sich für mich und auch die anderen als schwierig und oftmals mehrdeutig: Manche Antworten sind kurz und knapp, andere ausführlich und reich an Informationen über die Teilnehmerinnen. In beiden Fällen ist die Zuordnung schwierig. Manchmal sind zu wenig Informationen vorhanden, um eine befriedigende Entscheidung zu treffen. Oder die Antworten zeichnen sich durch einen so hohen Einfallsreichtum aus, dass es unangebracht scheint, die Teilnehmerin auf eine stereotype Kategorie zu reduzieren.

Bei der Präsentation der Resultate bedankt sich Simon, dass wir 600 solcher Zuordnungen gemacht hätten. Er witzelt, dass wir wohl mittlerweile eine Selbsthilfegruppe gegründet hätten. Damit spricht er ein für Earlybird wichtiges Problem an: Die Zuordnungen händisch vorzunehmen, braucht sehr viel Zeit und ist entsprechend teuer: Es »skalier« nicht und sei »bloody boring«, daher sei auch ihre Ambition, eine Software zu machen, die eine solche Zuordnung automatisiert. Wie schwierig und umstritten die Zuordnungen sind, wird gleich nochmals deutlich, als Simon erläutert, dass Konsens über ein »majority voting« eruiert wurde. Er gibt ein Beispiel einer Teilnehmerin, welche die jordanische Stadt Petra besuchen, am Strand die Seele baumeln lassen, mit Walhaien schwimmen und am Burning-Man-Festival in den USA teilnehmen möchte (siehe oben). Zwei der Experten oder Expertinnen, deren Identitäten anonymisiert wurden, hätten bei dieser Antwort gesagt, sie sei »progressiv postmodern«, einer »Hedonist«. Gemäß »majority voting« werde die Antwort dem Milieu »progressiv postmodern« zugeordnet, auch wenn das etwas »hemdsärmelig« sei. Mehr Experten seien besser.

Sabina kommentiert den vorliegenden Fall: Sie meint, diese Antwort sei extrem schwierig zuzuordnen, da die drei Antworten sehr unterschiedlich seien. Simon findet interessant, wie sowohl der Reiseort als auch die Antwort in die Beurteilung eingingen. Aber auch die Formulierungen seien interessant, sage ich. Zum Beispiel töne »die Seele baumeln lassen« etwas klischiert. Es

geht kleine Diskussion los, wie der Fall interpretiert werden soll. Nik ist für Young Performer, Beni wirft Action Sportler ein, was Sabina verneint. Beni verteidigt seine Antwort: Mit Walhaien schwimmen sei mit Risiko und Nervenkitzel verbunden! Nik meint »progressiv postmodern«. Simon beendet die Diskussion: man könne sehr lange über einen Fall »philosophieren« und Anna, Markus und Sabina hätten 600 zugeordnet. Bei 80 der 600 Antworten seien wir drei einer Meinung gewesen. Bei zirka 300 ließ sich mit »majority voting« ein Milieu zuordnen. Simon ist mit diesen Übereinstimmungen zufrieden und meint, dass sich damit etwas machen lasse.

-
- 4 »To attempt to ›repair‹ indexicality is to try to fix a machinery that isn't broken. Indexical expressions do present particular problems for programs in machine translation, survey analysis, linguistics, or other academic or practical endeavors for devising formal representations of practical actions. But once we no longer assume the classic posture of an objective observer, the general problem of indexicality dissolves. Ambiguities and misunderstandings sometimes arise in ordinary communicational activities, but even when they do they are repaired through further use of indexical expressions« (Lynch, 1992, S. 285).
 - 5 TF-IDF ist das »blunt instrument« in der Schublade der Data Science, wie Simon meint. Beel et al. (2016) zeigen, dass 83 Prozent aller textbasierten Empfehlungssysteme TF-IDF benutzen. Die Idee der »term specificity« – ein Wort ist für einen Text umso charakteristischer, in je weniger anderen Texten des Korpus es vorkommt – wurde von Karen Spärck Jones eingeführt (1972, siehe auch: Rieder 2020).
 - 6 Menschliche und algorithmische »Interpretation« sind insofern ähnlich, als beide relativ opak sind (Burrell, 2016).
 - 7 Verschiedene Studien weisen darauf hin, dass vor allem besser gebildete, weiße Männer aktiv an Wikipedia mitarbeiten (Hargittai & Shaw, 2015).
 - 8 »20 Minuten« ist die meistgelesene Schweizer Tageszeitung. Sie liegt an Bahnhöfen und Bushaltestellen kostenlos auf.
 - 9 Falls der höchste Wert einen bestimmten Schwellenwert nicht übersteigt (konkret: das 10-Prozent-Quantil in der Verteilung der höchsten Similaritäten), wird die Antwort der Kategorie »special groups« zugeordnet.
 - 10 Oder alternativ: Wer einen bestimmten Schwellenwert für ein Segment übersteigt, wird als »Hedonist« und »Young Performer« behandelt. Wie Martin Degeling betont, ist es von Vorteil, über mehr potenzielle Interessen Bescheid zu wissen: »Die Ungenauigkeit im Profiling macht aus der Perspektive der WerbevermarkterInnen allerdings durchaus Sinn. Es geht eben nicht darum, eine umfängliche und korrekte Persönlichkeitsbeschreibung vorzunehmen, sondern darum möglichst viele aktuelle Interessen zu kennen, für welche dann Werbung angezeigt werden kann« (Degeling, 2017, S. 24).

8.3 Die Operationen des algorithmischen Vergleichs

Am Beispiel von Earlybird lässt sich beobachten, wie bestehende Marketingsegmente (und anderes) in die Funktionsweise algorithmischer Kategorisierung eingefaltet werden. Um die Nutzerinnen den jeweiligen Kategorien zuzuordnen beziehungsweise zu bestimmen, welche Kategorien am besten passen, ist relativ unauffällig auch ein Vergleich eingefaltet.

Es scheint intuitiv klar zu sein, dass Algorithmen vergleichen. In Auseinandersetzung mit dem analytischen Vergleichsbegriff von Bettina Heintz argumentiere ich aber, dass sich die Funktionsweise des algorithmischen Vergleichs von »normalen« Vergleichen in zwei Hinsichten unterscheidet. Erstens finden algorithmische Vergleiche auf der Hinterbühne statt. Auf der personalisierten Benutzeroberfläche präsentieren Unternehmen nicht Vergleiche, sondern einfache Listen oder Empfehlungen, deren Ordnungsweise sich für die Nutzerin nicht ohne Weiteres erschließt. Zweitens zeigt ein detaillierter Blick auf die Operationen des algorithmischen Vergleichs, dass nicht Nutzerinnen oder Dinge verglichen werden, sondern paarweise Relationen. Scores beobachten nicht einzelne Nutzerinnen, sondern paarweise Kombinationen von Nutzerinnen und Dingen oder »Kategorien« wie Jugendmilieus.

Bettina Heintz (2010; 2016) stellt einen soziologischen Vergleichsbegriff zur Verfügung, der drei Aspekte des Vergleichs analytisch voneinander unterscheidet:

Erstens können Vergleiche nur dort durchgeführt werden, wo die Vergleichseinheiten als vergleichbar gelten. Vergleiche beruhen also auf einer Kategorisierung und können zur Grundlage weiterer Kategorienbildung werden. Im letzten Kapitel haben wir gesehen, dass im Bereich der Personendaten Vergleichbarkeit einerseits durch Encoding, d.h. Standardisierung möglicher Verhaltensweisen, und andererseits durch Vektorisierung hergestellt

11 Auch Gabrilovich und Markovitch vergleichen ihre Resultate mit einem »menschlichen Goldstandard«. »Humans have an innate ability to judge semantic relatedness of texts. Human judgements on a reference set of text pairs can thus be considered correct by definition, a kind of »gold standard« against which computer algorithms are evaluated« (Gabrilovich & Markovitch, 2007, S. 1609). Wie Menschen die »semantic relatedness« von Wörtern beurteilen wird über ein Korrelationsmaß damit in Beziehung gesetzt, wie Algorithmen die »semantic relatedness« berechnen. ESA erzielt eine höhere Korrelation mit »human judgement« als andere Algorithmen (Gabrilovich & Markovitch, 2007, S. 1609).

werden kann. Heintz macht darauf aufmerksam, dass die Grenzen des Vergleichs nicht einfach gegeben sind, sondern das Resultat gesellschaftlicher Institutionalisierung: Personen für die Volkszählung zu zählen ergibt erst dann Sinn, wenn die Bettlerin und die Königin nicht mehr als grundsätzlich verschiedene Wesen gelten, sondern beide gleichwertig der Kategorie »Mensch« zugerechnet werden können. Gleichzeitig hat der Vergleich auch Effekte: Werden zwei Dinge wiederholt über den Vergleich zueinander in Beziehung gesetzt, kann es zu einer Plausibilisierung ihrer Gleichheit kommen – zum Beispiel technische Hochschulen und Universitäten oder private und öffentliche Universitäten, deren Vergleichbarkeit von Universitätsrankings unterstellt wird.

Zweitens benötigt jeder Vergleich ein Vergleichskriterium und dazugehörige Verfahren, um Ähnlichkeiten und Differenzen zwischen den Vergleichseinheiten fest- beziehungsweise herzustellen: zum Beispiel die von Fourcade & Healy (2017b) beschriebenen Credit-Scores. Potenzielle Schuldner müssen dazu in Bezug auf ihre Bonität vermessen oder eingestuft werden, zum Beispiel anhand ihres Einkommens. Während zählen in manchen Fällen ausreicht, ist es meist komplizierter, insbesondere dann, wenn sich das interessierende Kriterium nicht direkt beobachten oder ohne Weiteres erschließen lässt (Mayntz, 2017) – zum Beispiel in der Frage, wie die »Qualität« von Kommentaren gemessen werden kann (siehe Kapitel 10.1). Es müssen dann Entscheidungen für bestimmte »proxies« getroffen werden, die beobachtet oder gemessen werden können und als Stellvertreter Auskunft über das gesuchte Vergleichskriterium geben (zum Beispiel quantitativ feststellbare Betreuungsverhältnisse an Universitäten als Stellvertreter für die Qualität der Lehre).

Drittens funktionieren Vergleiche als »Relationierungsinstrumente«:

Indem [Vergleiche] Sachverhalte anhand einer dritten Grösse auf ihre Unterschiede hin beobachten, stellen sie zwischen diesen einen Sinnzusammenhang her, der um einiges komplexer ist als die Ordnungsstruktur, die sich aus der blossen Zuordnung zur gleichen Kategorie ergibt. (Heintz, 2016, S. 307)

Dabei handelt es sich um kategoriale oder ordinale Relationen (siehe Fourcade 2016b; Heintz 2019 zu Ranglisten als Sonderform des Vergleichs), die Sinnzusammenhänge zwischen Einheiten herstellen, ohne dass diese sich explizit aufeinander zu beziehen brauchen: Es kommt zu Interdependenzen ohne faktische Vernetzung.

Algorithmische Vergleiche *relationieren* ihre Vergleichseinheiten auf unterschiedliche Weise. In »normalen« Vergleichen ist Relationierung ein Effekt des *öffentlichen* Vergleichs: Beispielsweise erzeugen Hochschul- oder Städte-rankings potenziell Effekte, indem sie zwischen den verglichenen Einheiten Sinnzusammenhänge herstellen und so einen Raum relevanter anderer Einheiten als Orientierungs- oder Abgrenzungsgrößen aufspannen: Die Zweitplatzierte kann die Verhaltensweisen der Erstplatzierten imitieren, muss aber gleichzeitig auch die aufholenden Konkurrentinnen im Blick behalten (siehe Espeland & Sauder 2007 für Law-School-Rankings). Der algorithmische Vergleich verfährt anders: Auch hier wird verglichen. Doch der Vergleich findet auf der Hinterbühne statt und wird unsichtbar gemacht. »Komunikativ erscheinen Empfehlungen nicht als Vergleiche (von Nutzerinnen), sondern als Relationen von Dingen« (Unternährer, 2020b, S. 385).

Bei normalen Vergleichen ist Relationierung ein Effekt der *Darstellung*. Die Vergleichseinheiten (zum Beispiel Skirennfahrerinnen) erfüllen die kategorialen Gleichheitsbedingungen, die einen Vergleich sinnvoll erscheinen lassen (i.e. weibliches Geschlecht, Qualifikation). Sie werden in Bezug auf ein Kriterium verglichen (i.e. wie schnell sie den Berg hinunterfahren) und in eine (ordinale) Vergleichsordnung gebracht (i.e. die Rangliste). Die einzelnen Skifahrerinnen werden in besser-schlechter-Relationen zueinander gesetzt. Die Positionierung in der Rangliste bestimmt, wen die Skifahrerinnen und ihre Zuschauerinnen als direkte Konkurrenz wahrnehmen.

Bei algorithmischen Vergleichen ist Relationierung eine rechnerische Operation, die Einheiten gleichen Typs (Nutzerinnen mit anderen Nutzerinnen), aber auch Einheiten unterschiedlichen Typs (Nutzerinnen mit Dingen wie beispielsweise Filmen) paarweise zueinander in Beziehung setzt (zum Beispiel die Berechnung einer Korrelation oder einer Distanz). Scoring- oder Empfehlungsalgorithmen relationieren vor allem durch die Feststellung von Differenz und Ähnlichkeit, d.h. während der Durchführung des Vergleichs. Sind die Vergleichseinheiten vergleichbar (zum Beispiel Nutzerinnen eines Collaborative-Filtering-Empfehlungssystems; siehe Kapitel 10.4), wird jede Einheit (zum Beispiel Nutzerinnen) mit jeder anderen paarweise in Beziehung gesetzt. Dann werden die Relationen zwischen den Nutzerinnen in Bezug auf »Ähnlichkeit« verglichen. Hinsichtlich *einer spezifischen* Nutzerin lassen sich dann die ähnlichsten anderen Nutzerinnen bestimmen und als »genügend gleich« kategorisieren.

Die Differenz zu normalen Vergleichen besteht darin, dass nicht einzelne Einheiten verglichen werden, sondern die Relationen zwischen Einheiten.

Während zum Beispiel bei Leistungsvergleichen die Leistungsvermessung einer Einheit relativ unabhängig von den Leistungen anderer Einheiten ist, ist die Berechnung einer Empfehlung auf der Grundlage der Daten nur einer Nutzerin nicht möglich.¹² »[A]s data becomes meaningful and useful only when entries from different sources (users) are combined and analysed jointly« (Charitsis et al., 2018, S. 828). Der algorithmische Vergleich erzeugt neue Relationen durch eine relationale Verdichtung: Alle Einheiten werden zu allen anderen Einheiten in Beziehung gesetzt und für jede Einheit wird dann verglichen, welche anderen Einheiten am ähnlichsten sind.

Algorithmische Vergleiche relationieren ihre Einheiten über paarweise »matchings«. Das hat Konsequenzen. Der Vergleich neuer digitaler Technologien mit der amtlichen Statistik legt nahe, dass sich ein grundlegender Wandel vollzieht. Dieser verändert nicht nur, *wann* kategorisiert wird, sondern auch *wie* das geschieht (Heintz, 2021). Alain Desrosières beschreibt in seiner Geschichte der großen Zahlen, dass Kategorien der Messung vorausgehen: »conventions of equivalence, encoding, and classification [which] precede statistical objectification« (Desrosières, 1998, S. 236). Algorithmische Vergleiche drehen diese Abfolge tendenziell um:

the relationship between classification and measurement is turned on its head. At least to some extent, datafication dissolves the importance of classification in measurement (in Desrosières's sense). The relationship between classification and measurement is turned upside down. Contrary to what Desrosières (1998) stated twenty years ago, classification does not precede measurement but becomes a result of it. (Mennicken & Espeland, 2019, S. 237)

Erst wird gerechnet, dann wird kategorisiert.¹³ Diese Umkehrung basiert auf einer neuen Art der Datenerzeugung, die zwar nicht auf klassischen Kategorien basiert, aber auf quasi-kategorialen, encodierten Verhaltensweisen (siehe Kapitel 7). Beruhte beispielsweise die Volkszählung zuerst auf einer

12 Außer bei inhaltsbasierten Empfehlungen. Aber auch bei inhaltsbasierten Empfehlungen stehen die Dinge, d.h. die Inhalte, nicht für sich, sondern immer nur im Verhältnis zu allen anderen Inhalten (siehe Kapitel 10.3).

13 Natürlich ist es nicht so einfach: Encoding lässt sich im Prinzip als basale Form der Kategorisierung verstehen. Und es ist auch nicht davon auszugehen, dass Kategorisierung eine ungeordnete Wirklichkeit ordnet, sondern schon bestehende Formen von Ordnung neu sortiert.

Fremdzuschreibung der Gezählten zu verschiedenen Klassifikationen wie Geschlecht, Beruf, Religion, etc., dann auf einer Selbstzuordnung, bei der Kreuze in entsprechenden Feldern gemacht wurden, so setzen aktuelle Kategorisierungstechnologien auf Verhaltensdaten. Anstatt die Population der Internetnutzerinnen direkt nach ihren Identitätskategorien oder nach ihren Interessen zu befragen, lassen sich digitale Daten als Spuren dessen lesen, was die Nutzerinnen »tatsächlich« tun (Cardon 2017, kritisch: Rouvroy 2013, affirmativ: Pentland 2014). Statt von kategorialen Gemeinsamkeiten der Subjekte auf ihr Verhalten zu schließen, wird von Verhaltensregelmäßigkeiten auf quasi-kategoriale Gemeinsamkeiten geschlossen (siehe insbesondere: Cheney-Lippold 2017; Fisher & Mehozay 2019).

Personenkategorien und Passungsverhältnisse zwischen Nutzerinnen und Dingen werden nicht im Voraus bestimmt, sondern abgeleitet. Personen werden dabei nicht als Essenzen verstanden, sondern als Gesamtheit ihrer Verhaltensweisen in Relation zu den Verhaltensweisen aller anderen Nutzerinnen. Algorithmische Vergleiche basieren dementsprechend auf einer »ontology of association« (Amoore, 2011, S. 27). Der algorithmische Vergleich ist nicht daran interessiert, zeitlose Wahrheiten über die Nutzerinnen zu produzieren, sondern Personendaten »actionable« zu machen, d.h. neue Relationen daraus abzuleiten: »[The data derivative] is not centred on who we are, nor even on what our data says about us, but on what can be imagined and inferred who we might be« (Amoore, 2011, S. 28). Manche Relationen zwischen Nutzerinnen und Dingen erscheinen vielversprechender als andere. Dafür braucht es einen Vergleich, der für jede Nutzerin die passendsten Matches eruieren kann.

8.4 Matching Devices

Der Wert von Daten besteht nicht in den Daten per se, sondern in dem, was aus den Daten abgeleitet werden kann: »good matches«, d.h. (neue) Relationen zwischen Nutzerinnen und Dingen, welche die Relation von Nutzerin und Unternehmen festigen. Dementsprechend sind auch nicht Daten selbst das veräußerbare Produkt, sondern »tools«, die »good matches« generieren können.

Although data brokers are fuelled with individual information from various sources, they do not sell personal data themselves, but business-oriented in-

formational products based on specific calculations performed on personal data. Their products are lists of qualified leads for direct marketing, marketing segments describing lifestyles and consumption patterns in general terms (e.g. Personixx segmentation by Acxiom), enrichment of existing databases, i.e. adding variables to their customers' CRM bases, and risk calculation (scoring) on individuals or profiles, particularly in the banking and insurance sectors. (Beauvisage & Mellet, 2020, S. 85)

Personendaten werden in »matching devices« (siehe Karpik 2010 zu »judgment devices«, Muniesa et al. 2007 zu »market devices«) inkorporiert, welche die ökonomischen Beziehungen zwischen Plattform-Unternehmen, Nutzerinnen und Dritten organisieren und am Laufen halten.¹⁴

Von Personendaten zum Classifier-Tool

Meine Feldforschung bei Earlybird endet, kurz bevor Urs das Kategorisierungsexperiment an einer akademischen Konferenz präsentiert. Einige Monate später erklärt mir Nik in einer Nachbesprechung, dass sie den »classifier« zu einem Tool weiterentwickelt hätten, das nicht nur bei Earlybird, sondern auch bei anderen Partnerunternehmen eingesetzt werden kann. Earlybirds Marketingsegmente wurden fallengelassen. Nutzerinnen werden nun mit Hilfe von Clusteranalysen direkt miteinander in Beziehung gesetzt, d.h. Nutzerinnen, deren Freitexte sich im Wikipedia-Vektorraum nahe beisammen befinden, werden derselben Kategorie zugeordnet. Dieses »smart customer segmentation«-Tool kann auch in ganz anderen Kontexten eingesetzt werden, da der Bezug zum ursprünglichen Wettbewerb und den dadurch generierten Daten praktisch unsichtbar gemacht wurde.

Nik beginnt seine Erklärung mit der Aussage, dass sie die Jugendmilieus abgeschafft hätten. Diese seien schlecht und »Kacke«. Die Milieubeschreibungen hätten nichts ausgesagt. Sie würden es nun anders machen, näm-

- 14 Nicht nur Unternehmen befassen sich mit der Erzeugung von »good matches«: Auch Plattformen wie Amazon, Youtube, Instagram, etc. sind auch die Nutzerinnen damit beschäftigt, ihre Produkte, Videos oder Bilder so zu gestalten, dass sie für ihre Follower möglichst ansprechend sind. Sie optimieren ihre Posts und deren Verbreitung, indem sie die algorithmische Selektion und Priorisierung anhand von Alltagstheorien (»folk theories« bei Eslami et al. 2016 oder »algorithmic gossip« bei Bishop 2019) antizipieren – beispielsweise indem sie ihre Inhalte in spezifischen Zeitfenstern teilen, wenn möglichst viele ihrer Follower wahrscheinlich online sind.

lich mit Clusteringverfahren. Nik zeichnet verschiedene mögliche Wettbewerbsantworten auf: Berg, Pilatus, Macchu Pichu. Diese Antworten verorten sie dann im Wikipedia Korpus (angereichert durch das »20 Minuten«-Korpus für den Schweizer Sprachgebrauch, zum Beispiel Begriffe wie »Natel«, »Velo«, etc.) und drücken jeden Begriff der Antwort als Vektor aus. Wenn man jetzt »bergaffine Leute« heraussuchen möchte: Wie bringt man das der Maschine bei? Man müsse wissen, wie ähnlich die Wörter zueinander seien. Das macht der Algorithmus. Die Arbeit des Algorithmus stellt er als einen kreisförmigen Pfeil dar: Darin würden die Antworten so verarbeitet, dass Gruppen wie die Marketingsegmente herauskommen, nur dass sie nicht mehr die Segmente verwenden und auf die Antworten matchen, sondern direkt aus den Antworten ähnliche Gruppen generieren.

Es ließen sich dann aus den Daten verschiedene »Personas« definieren. Personasseien Vorstellungen der Firma oder der Marketingabteilung darüber, wer ihre Kunden »da draussen« sind. Er gibt ein Beispiel: Rudi ist im Sportverein, rudert, hat 10 kg Übergewicht. Seine Kunden würden sich ihre Kunden in solchen Stereotypen vorstellen. Diese Analysen würden es ermöglichen, neue und genauere Personas zu definieren oder zu überprüfen, ob bestehende Personas sich auch tatsächlich in den Daten wiederfinden lassen. Man könne den Stereotyp Rudi mit seinen tatsächlichen Interessen vergleichen.

Man müsse sich dann entscheiden, wie viele Zielgruppen man brauche und überhaupt »handlen« könne, d.h. für wie viele Zielgruppen die Marketingabteilung Kampagnen entwickeln könne. Die Kampagnen basieren dann nicht mehr auf Segmenten, sondern auf den meistverwendeten oder wichtigsten Begriffen der einzelnen Cluster. Diese Begriffe ließen sich »ausspielen«. Man erhalte also »interessensbasierte« Cluster, die man für Marketingkampagnen verwenden könne.

Nik nutzt den Classifier auch für die Kundenakquise, d.h. als Demonstration, um anderen Unternehmen Datenanalyse als Dienstleistung schmackhaft zu machen. Das Ziel des Tools sei es, Firmen dabei zu helfen, ihre Kunden dank Textdaten besser kennen zu lernen und ihnen den »Mehrwert« von Kundendaten zu vermitteln. Er konnte auch schon den ersten Erfolg verbuchen: Ein weiteres Reiseunternehmen wollte den Classifier ausprobieren, um mehr über seine eigenen Kundinnen zu erfahren. Der Classifier als »freebie« soll als Fuß in der Tür funktionieren: Potenzielle Kundinnen sollen vom Potenzial

der Datenanalyse überzeugt werden, damit sie Earlybird Digital mit weiteren Analysen beauftragen.

Aus Earlybirds Kategorisierungsexperiment entsteht ein Classifier-Tool, das in zweifachem Sinn ein Beziehungsgenerator ist. Das Tool generiert prospektive Beziehungen zwischen Nutzerinnen und Dingen (beziehungsweise »Deals«, die Partnerunternehmen anbieten). Aus bestehenden Nutzerdaten leitet der Classifier kategoriale Zugehörigkeiten ab, die darüber Auskunft geben sollen, welche weiteren Dinge für die Member interessant sein könnten. Die Segmente sollen zu einem Instrument für die Marketingabteilung werden, um in ihren verschiedenen Kampagnen den einzelnen Mitgliedern die »richtigen«, für sie interessanten, »Deals« zu präsentieren.

Andererseits lassen sich aus Personendaten nicht nur prospektive Relationen zwischen dem datengenerierenden Unternehmen und seinen Nutzerinnen ableiten. Auch Relationen zu Dritten lassen sich generieren und festigen. Earlybirds Idee ist es, zu einer Werbeplattform zu werden, die den Partnerunternehmen Zugang zu spezifischen Segmenten innerhalb der Zielgruppe aller Jugendlichen bieten kann.¹⁵ In diesem Sinn werden Personendaten – d.h. deren Generierung, Verarbeitung und In-Wert-Setzung – zur zentralen Ressource in der Generierung von Relationen zwischen der Plattform, den Werbekundinnen und den Nutzerinnen (Fourcade & Kluttz, 2020; Gerlitz & Helmond, 2013). Die durch Personendaten ermöglichten Relationierungsweisen dienen als Kitt, der Unternehmen und Nutzerinnen (und Dritte) über »good matches« von Nutzerinnen und Dingen zusammenhalten soll.

»Good matches«, die auf Personendaten beruhen, ermöglichen ökonomische Beziehungen. In dieser Herstellung von Passungsverhältnissen zwischen verschiedenen ökonomischen Akteuren (Plattformen, Nutzerinnen, (Werbe-)Kundinnen) sehe ich das zentrale Wertversprechen der digitalen Ökonomie, deren Geschäftsmodelle vor allem aus Werbung, Kunden- und Nutzerbindung besteht. Digitale Verhaltensdaten informieren über vergangene Handlungen von Nutzerinnen. Der Wert dieser Daten kommt aber erst dann zu Stande, wenn die Daten zu anderen Daten algorithmisch in Relation gesetzt werden, um daraus neue, potenziell daten-

15 Earlybird ist bereits eine Plattform, die Partnerunternehmen und Banken Zugang zur »Zielgruppe der Jugendlichen« als Ganzes bietet.

und profitgenerierende, Relationen abzuleiten. Dies führt wiederum zu neuen relationalen Verwicklungen und Personendaten.

Die Generierung *neuer* Relationen ist entscheidend und wird beispielsweise in der prominenten Kritik von Pariser (2011) übersehen. Nutzerinnen befinden sich nicht in einem »you-loop« (Pariser, 2011, S. 16): »your profile is only made meaningful and commodifiable to marketers in and alongside the context of other users' profiles« (Kant, 2020, S. 35).¹⁶

Dass der Classifier von Earlybird Digital am Ende selbst zu einem Geschenk wird, illustriert die Relevanz von Geschenken in der digitalen Ökonomie. Personendaten werden als beziehungsgenerierende Geschenke »hervorgehoben«. Die Beziehungen werden kurzzeitig entfernt, um Personendaten als warenförmige Ressource weiter verarbeiten zu können, um daraus neue Relationen abzuleiten. In der Form passender Empfehlungen, dem »richtigen Angebot zur richtigen Zeit« oder personalisierter Werbung, kehren Personendaten zu den Nutzerinnen zurück.¹⁷

16 Siehe auch Wendy Chun: »[V]alue is not generated by one YOU but rather by a plethora of YOUS: by the very interconnections between the various YOUS« (Chun, 2016, S. 118).

17 Minna Ruckenstein und Julia Granroth (2019) ziehen gar in Betracht, dass Unternehmen und Nutzerinnen ein »intimes« Verhältnis zueinander eingehen. Die »intimacy of surveillance« kann auf Seiten der Nutzerinnen durchaus lustvoll sein – wenn sie von den Unternehmen richtig erkannt und eingeschätzt. Aber auch negative Erfahrungen der Kategorisierung gehören zum Alltag von Nutzerinnen, wenn sie sich falsch kategorisiert werden oder eine (zu gute) Empfehlung ihnen das Gefühl gibt, in ihren heimlichen Interessen ertappt worden zu sein.

